

IN THE CLAIMS:

The text of all pending claims, (including withdrawn claims) is set forth below. Cancelled and not entered claims are indicated with claim number and status only. The claims as listed below show added text with underlining and deleted text with ~~striketrough~~. The status of each claim is indicated with one of (original), (currently amended), (cancelled), (withdrawn), (new), (previously presented), or (not entered).

Please CANCEL claim 21, AMEND claims 1 and 2 and ADD new claim 23 in accordance with the following:

1. (currently amended) A method for collecting documents linked to each other from a network by crawling the network, comprising:

collecting a number of documents into a collected document group until the number of documents is equal to or larger than a predetermined value, the documents being collected from inside a community through the network based on a reference of each document in the collected document group, the reference defining a relationship between referring documents and referenced documents inside or outside the community which are linked to the referring documents; ~~and~~

collecting a newly collected document group from inside and outside the community based on the reference of the collected document group from inside the community; and

indicating a level of significance of a collected document belonging to the newly collected document group according to references to the collected document and information about a position of the collected document in the network, with the significance of the collected document increasing with an increased number of references to the collected document from the referring documents thereof at network locations having low similarity to the position of the collected document in the network.

2. (currently amended) The method according to claim 1, further comprising: ~~computing a significance level indicating a level of significance of a collected document belonging to the newly collected document group according to references to the collected document, and information about a position of the collected document in the network; and~~

determining new documents to be collected based on the references from and the significance level of the collected document.

3. (previously presented) The method according to claim 2, wherein the new documents are determined separately for inside the community and for outside the community.

4. (previously presented) The method according to claim 3, further comprising:
presenting a result of retrieving the new documents separately for inside the community and outside the community.

5. (previously presented) The method according to claim 2, further comprising:
determining whether any of the documents is in the community according to information indicating the position of the document in the network.

6. (cancelled)

7. (currently amended) A method for collecting documents linked to each other from a network by crawling the network, comprising:

providing a positive sample document group relating to a field, and a negative sample document group related to the field;

determining a prospective document for collection that is related to the field based on a reference relation to the positive sample document group and the negative sample document group by computing a reference score indicating a level at which the prospective document is referenced only by at least one document in the positive sample document group based on the reference relation which defines a relationship between original documents and linked documents belonging to the positive sample document group or to the negative sample document group which are linked to the original documents; and

collecting the prospective document having a high reference score; and

indicating a level of significance of the prospective document according to references to the prospective document and information about a location of the prospective document in the network, with the significance of the prospective document increasing with an increased number of references to the prospective document from referring documents at network locations having low similarity to the location of the prospective document in the network.

8. (currently amended) A method for collecting documents linked to each other from a network by crawling the network, comprising:

providing a positive sample document group relating to a field, and a negative sample document group less related to the field;

determining a prospective document for collection that is related to the field based on a reference relation to the positive sample document group and the negative sample document group by computing a co-reference score indicating a level at which the prospective document is referenced together with a first collected document in the positive sample document group for a second collected document referenced by a third collected document referring to a fourth collected document in the positive sample document group based on the reference relation which defines a relationship between original documents and linked documents belonging to the positive sample document group or to the negative sample document group which are linked to the original documents; ~~and~~

collecting the prospective document having a high co-reference score ; and
indicating a level of significance of the prospective document according to references to the prospective document and information about a location of the prospective document in the network, with the significance of the prospective document increasing with an increased number of references to the prospective document from referring documents at network locations having low similarity to the location of the prospective document in the network.

9. (cancelled)

10. (previously presented) The method according to claim 1, further comprising:
summarizing the collected documents based on a referencing expression used in the collected documents.

11. (previously presented) The method according to claim 1, further comprising:
assigning a keyword to the collected documents based on a referencing expression used in the collected documents.

12. (original) The method according to claim 1, further comprising:
not assigning a keyword based on the referring expression when the referencing expression is used regardless of a content of a referenced document.

13. (original) The method according to claim 11, further comprising:
counting a number of different documents referenced using the referencing expression;
and

not assigning the keyword based on the referencing expression when the number of
different documents is equal to or larger than a predetermined value.

14. (original) The method according to claim 11, further comprising:
counting a reference frequency at which each collected document is referenced by the
referencing expression when the number of different documents is smaller than a predetermined
value; and

determining whether or not the referencing expression is assigned as the keyword based
on the number of different documents and the reference frequency.

15. (original) The method according to claim 11, further comprising:
combining the keyword based on the referencing expression with a keyword extracted
from text of the collected document, and a keyword extracted from information indicating a
position in the network about the collected document.

16. (currently amended) A method for retrieving documents from a terminal belonging to
a community in a network, comprising:

transmitting information for retrieval of the documents to a server; and
receiving the documents retrieved separately from inside and outside the community
according to the information for retrieval together with information indicating a significance level
for the community, the significance level indicating an importance of a document based on
information about references in a collected document group, and information indicating a
position of the document, with the significance level of the document increasing with an
increased number of references to the document from referring documents at network locations
having low similarity to the position of the document in the network.

17. (currently amended) A document collection apparatus collecting documents linked to
each other from a network by crawling the network, comprising:

a next prospect determination unit determining a prospect to be collected next based on
a reference relation of a collected document, the reference relation defining a relationship

between original documents and other documents inside or outside the community which are linked to the original documents;

a community determination unit determining whether or not the prospect is in a community in the network according to information indicating a position in the network of the prospect; and

a document collection unit collecting the prospect ~~from the network, wherein said document collection unit collects the prospect and other prospects~~ from inside and outside the community after collecting documents larger in number than a predetermined value from inside the community, and computing a level of significance of the prospect according to references to the prospect and the information about the position of the prospect in the network, with the significance of the prospect increasing with an increased number of the references to the prospect from referring documents at network locations having low similarity to the position of the prospect in the network.

18. (currently amended) A document collection apparatus collecting documents linked to each other from a network by crawling the network, comprising:

a next prospect determination unit determining a prospect to be collected next based on a reference relation between a positive sample document group which is a document group related to a field and a negative sample document group which is a document group less related to the field; and

a document collection unit collecting the prospect from the network; wherein the reference relation is a relationship between original documents and other documents belonging to the positive sample document group or to the negative sample document group which are linked to the original documents, and computing a level of significance of the prospect according to references to the prospect and information about a position of the prospect in the network, with the significance of the prospect increasing with an increased number of the references to the prospect from referring documents at network locations having low similarity to the position of the prospect in the network.

19. (currently amended) A computer-readable recording medium recording a program used to direct a computer to control collection of documents linked to each other from a network by crawling the network, comprising:

collecting a number of documents into a collected document group until the number of documents is equal to or larger than a predetermined value, the documents being collected from

a community through the network based on a reference of each document in the collected document group, the reference defining a relationship between referring documents and referenced documents inside or outside the community which are linked to the referring documents; ~~and~~

collecting a newly collected document group from inside and outside the community based on the reference of the collected document group from inside the community; and

indicating a level of significance of a collected document belonging to the newly collected document group according to references to the collected document and information about a position of the collected document in the network, with the significance of the collected document increasing with an increased number of references to the collected document from the referring documents thereof at network locations having low similarity to the position of the collected document in the network.

20. (currently amended) A computer-readable recording medium recording a program used to direct a computer to control collection of documents linked to each other from a network by crawling the network, comprising:

providing a positive sample document group which is a document group relating to a field, and a negative sample document group which is a document group less related to the field;

determining a document to be collected relating to the field based on a reference relation to the positive sample document group and the negative sample document group, the reference relation defining a relationship between original documents and other documents inside or outside the community which are linked to the original documents; ~~and~~

collecting the document to be collected from the network; and
indicating a level of significance of the document according to references to the document and information about a location of the document in the network, with the significance of the document increasing with an increased number of references to the document from referring documents at network locations having low similarity to the location of the document in the network.

21. (cancelled)

22. (currently amended) A method for collecting documents linked to each other from a network by crawling the network, comprising:

collecting a number of documents into a collected document group until the number of documents is equal to or larger than a predetermined value, the documents being collected from inside a community through the network based on references in the documents in the collected document group, each reference defining a relationship between a referring document and a referenced document, inside or outside the community, which is linked to the referring document;
~~and~~

collecting a newly collected document group from inside and outside the community based on the reference of the collected document group from inside the community; and

indicating a level of significance of a collected document belonging to the newly collected document group according to references to the collected document and information about a position of the collected document in the network, with the significance of the collected document increasing with an increased number of references to the collected document from referring documents at network locations having low similarity to the position of the collected document in the network.

23. (new) The method according to claim 1, wherein
a significance of a document referenced by a URL having high similarity is assumed to be low,

a document referenced by a larger number of documents are more significant, and a document referenced by a significant document having a low similarity of URL is significant, and

a similarity of a URL is defined according to character information about the URL such that a similarity can be a lowest when a document has different server address, path, and file name and such that a similarity of a document in a mirror site or a same server can be high.